Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

(2)

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION: Unclassified	1b RESTRICTIVE MARKINGS: NA
2a SECURITY CLASSIFICATION AUTHORITY: NA	3 DISTRIBUTION / AVAILABILITY OF REPORT: Distribution Unlimited
2b DECLASSIFICATION / DOWNGRADING SCHEDULE: NA	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): Rasc-3-5-25038	5 MONITORING ORGANIZATION REPORT NUMBER(S): NA
6a NAME OF PERFORMING ORGANIZATION: Scripps Clinic & Res. Fnd. / 6b OFFICE SYMBOL (If applicable): NA	7a NAME OF MONITORING ORGANIZATION: Office of Naval Research
6c ADDRESS (City, State, and ZIP Code): Res. Inst. of Scripps Clinic, MB5 10666 N. Torrey Pines Rd. La Jolla, CA 92037	7b ADDRESS (City, State, and ZIP Code): 800 N. Quincy St. Arlington, VA 22217-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION: Office of Naval Research / 8b OFFICE SYMBOL (If applicable): ONR	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-89-J-1174
8c ADDRESS (City, State, and ZIP Code): 800 N. Quincy St. Arlington, VA 22217-5000	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 61153N / PROJECT NO: RR04106 / TASK NO: 441n011-01 / WORK UNIT ACCESSION NO:

11 TITLE (Include Security Classification)
(U) Defining Protein Electrostatic Recognition Processes

12 PERSONAL AUTHOR(S)
Getzoff, Elizabeth D., Roberts, Victoria A.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Annual	FROM 12/88 TO 11/89	1989 November 30	7

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Electrostatic Recognition Macromolecular Complexes Computer Graphics
06	03		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Using a coupled computational and interactive computer graphic approach is used to we seek to elucidate the nature of electrostatic forces controlling the interaction of *stable*, *catalytic*, and *transient* binding complexes, which are basic to biological recognition processes. For each of these three kinds of binding, a specific macromolecular system has been selected for study. For the stable complex between the enzyme lysozyme and an antilysozyme antibody, ~~we will~~ determine the role of electrostatic forces in the specificity and efficiency of binding. For the catalytic binding of the R67 plasmid enzyme dihydrofolate reductase, ~~we will~~ one deduce the role of electrostatic forces in the functionally important binding of the dihydrofolate substrate and the NADPH cofactor. For the transient electron-transfer complex between plastocyanin and cytochrome c, we will predict critical residues and interaction area for the electron transfer interaction. (cont.)

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	(U)
22a NAME OF RESPONSIBLE INDIVIDUAL: M. Marron	22b TELEPHONE (Include Area Code): 202/696-4760 / 22c OFFICE SYMBOL: ONR

DD Form 1473, JUN 86 — Previous editions are obsolete — SECURITY CLASSIFICATION OF THIS PAGE (U)

S/N 0102-LF-014-6603

19.

The specific aims are to develop and apply 1) novel algorithms for rapid calculation of electrostatic interactions allowing local side chain mobility and conformational changes, and dealing explicitly with solvent and counter-ions 2) new types of models using interactive computer graphic representations made possible by coupling supercomputers to computer graphics, 3) comprehensive automatic docking methods based upon electrostatic fields, and 4) detailed characterization of mechanisms of functionally important electrostatic interaction for three protein systems, including testable predictions for each of the three general types of protein recognition. These predictions of molecular properties will enhance understanding of structure-function relationships, guide site-directed mutagenesis experiments, and aid in the design of inhibitors, and proteins and protein analogues with improved or new specific functions. Thus, these studies of protein electrostatic recognition processes will provide results relevant to many disease processes.

DATE: 30 November 1989

ANNUAL REPORT ON CONTRACT N00014-89-J-1174 R&T CODE 441n011

PRINCIPAL INVESTIGATOR: Elizabeth D. Getzoff

CONTRACTOR: Scripps Clinic and Research Foundation

CONTRACT TITLE: Defining Protein Electrostatic Recognition Processes

START DATE: 1 December 1988

RESEARCH OBJECTIVE: The design and application of a tightly coupled computational and interactive computer graphics approach to elucidate the nature of electrostatic forces controlling the interaction of three general classes of electrostatic recognition processes: stable, catalytic and transient binding complexes.

PROGRESS (Year 1):

Precollision Orientation by Electrostatic Forces. Attempts to define macromolecular complexes by docking static crystallographic structures may fail because local structural changes are not accommodated. We have approached the problem of protein-protein interaction by considering the relative orientations of the two molecules at a separation distance where the precollision orientation is dictated by long-range electrostatic forces.

To search for all possible favorable energy orientations of two molecules, two problems needed to be overcome. First, the search must be computationally efficient, and, second, the molecules must be maintained at a constant minimum distance. The first problem was solved by calculating an electrostatic grid about one molecule and using the grid to calculate the total electrostatic interaction with the other molecule in its many orientations. The second problem was solved by making and aligning expanded molecular dot surfaces for each molecule and checking these surfaces for interpenetration. The program TURNIP used these methods to carry out a systematic six dimensional search of paired molecular orientations evaluated by their resulting electrostatic energy.

In the first year of this contract, we have developed and optimized the TURNIP program, which consists of two parts. First, the electrostatic potential due to molecule 1, which remains fixed, was calculated onto a grid that was large enough to encompass all positioning of molecule 2 with its expanded surface. Second, molecule 2 was then moved about molecule 1 by systematically

matching one expanded surface point from each molecule along the intermolecular axis (a line between the centers of coordinates of the two molecules). For each matched pair of points, molecule 2 was rotated in 10° increments about the intermolecular axis. For each orientation, in which the expanded surfaces interpenetrated, molecule 2 was incrementally translated a small distance away from molecule 1 along the intermolecular axis and retested for penetration. For each position of molecule 2, the electrostatic potential energy was calculated as the sum of the charges of each atom of molecule 2 multiplied by the potential at its nearest grid point, which represented the electrostatic potential of molecule 1. The calculation of the electrostatic potential from molecule 1 at the grid points is proportional to the number of atoms of 1, while the calculation of the intermolecular energies is proportional to the number of atoms in molecule 2. Thus, the grid saves substantial calculation time compared with a pairwise calculation for each orientation, which would be proportional to the number of atoms in molecule 1 times the number of atoms in molecule 2. This calculation constitutes a full search over the six degrees of freedom maintaining a constant distance between the two molecules.

Because the program is divided into two parts, each part can be run on computers with specifications that match the task at hand. For a complete search involving two ~100-residue proteins, the energy grid is sufficiently large that it cannot be held in memory on our Cray XMP at Scripps Clinic. Instead, the electrostatic potential grid is created on the Cray and written to a file that is then transferred to our Convex C2 supercomputer where the rotation and translation of molecule 2 occurs. The program TURNIP is written in the programming language C, and has been optimized to take advantage of the capabilities of each machine. In the last year, the C compiler for the Cray has been greatly improved. Together with vectorization and optimization of the code, this has improved the speed by more than a factor of 20. Calculation of the electrostatic grid is currently done with a Coulomb potential and takes about an hour of CPU time on the Cray supercomputer to generate a grid with spacing of 0.5Å large enough to maintain a separation distance of 6Å between two ~100-residue proteins in all orientations. More complex electrostatic methods could easily be implemented and would run efficiently on this machine. The virtual memory of the Convex allows the entire grid to be read in at once. A complete search of all orientations of two ~100-residue proteins takes about 28 hours of CPU time. A coarser grid or a restricted rotational space can substantially reduce the time necessary for either part of the program.

We have applied TURNIP to the investigation of the transient interaction between two electron transfer proteins, plastocyanin and cytochrome *c*, and have compared these results with those

found by visual alignment of electrostatic fields displayed with computer graphics. The computer graphics method gives two favorable orientations that vary by a 180° rotation. The comprehensive computational search methods provide a family of orientations centered about those found by computer graphics and reveal that there are many similar energy orientations when the molecular surfaces are separated by 6 and 12Å. All orientations have the exposed heme edge of cytochrome *c* facing the acidic patch of plastocyanin. This region of favorable approach of cytochrome *c* is consistent with experimental data showing that electron transfer proteins may meet in multiple orientations and then translocate along the protein surfaces to reach an orientation appropriate for electron transfer.

Analysis of the U10 Mutant of the Phosphorylcholine-Binding Antibody S107. The U10 antibody results from a single-site mutation of heavy chain Asp 101 in the parent antibody, S107, to an alanine. The mutated residue is over 9 Å distant from the antigen binding site, but results in a complete loss of phosphorylcholine binding activity. A combination of molecular dynamics and minimization techniques suggested that the mutant antibody may fold so that Arg 94 of the heavy chain, which forms a salt-bridge with Asp 101 in a related phosphorylcholine-binding antibody, McPC603, extends into the antigen binding pocket to electrostatically and sterically block the binding of the hapten. This work has been published in the Proceedings of the National Academy of Science, USA (copy enclosed).

Interactive Computer Graphics. We have coupled the molecular dynamics programs AMBER and DISCOVER, running either on the Cray or the Convex supercomputers, through a network connection to our interactive graphics program **Flex** running on another computer. Now, dynamics or energy minimizations done on the supercomputers can be viewed interactively on Sun workstations while the calculations are being performed.

WORK PLAN: In the coming year, the methodical search program will be applied to the components of two protein systems of known structure, where electrostatics appears to play a major role in molecular recognition and orientation: a lysozyme/antibody complex and the tetramer of plasmid dihydrofolate reductase. We are also continuing development of methods for visualizing and comparing the suggested multiple orientations that result from these calculations.

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5532–5536, July 1989
Immunology

Significant structural and functional change of an antigen-binding site by a distant amino acid substitution: Proposal of a structural mechanism

(antibody structure/computer modeling/immunoglobulin sequence/phosphocholine binding/somatic mutation)

NADINE C. CHIEN*†, VICTORIA A. ROBERTS‡, ANGELA M. GIUSTI*, MATTHEW D. SCHARFF*, AND ELIZABETH D. GETZOFF‡

*Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461; and ‡Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037

Contributed by Matthew D. Scharff, March 9, 1989

ABSTRACT **To study the molecular basis for antibody diversity and the structural basis for antigen binding, we have characterized the loss of phosphocholine (*P*-Cho) binding both experimentally and computationally in U10, a somatic mutant of the antibody S107. Nucleotide sequencing of U10 shows a single base change in J_H1, substituting Asp-101 with Ala, over 9 Å distant from the *P*-Cho-binding pocket. Probing with anti-idiotypic antibodies suggests local, not global, conformational changes. Computational results support a specific structural mechanism for the loss of *P*-Cho binding. The U10 mutation eliminates the charged interaction between Asp-101 and Arg-94, which allows the Arg-94 side chain to disrupt *P*-Cho binding sterically and electrostatically by folding into the *P*-Cho-binding site. These results specifically show the importance of the Arg-94 to Asp-101 side chain salt bridge in the heavy-chain CDR3 conformation and suggest that residues distant from the binding site play an important role in antibody diversity and inducible complementarity.**

Somatic diversity of germ-line immunoglobulin variable (V) region genes during the course of the immune response can alter the affinity and specificity of antibodies (1–5). Correlation of individual amino acid substitutions with changes in antigen binding is often difficult because most monoclonal antibodies with significant changes in affinity differ by multiple amino acid substitutions in both heavy (H) and light (L) chains (1–6). To examine the impact of somatic mutation on V region structure, we have developed an *in vitro* assay to isolate spontaneously generated somatic mutants of anti-phosphocholine (*P*-Cho) antibodies with altered specificity and affinity (7). The S107.3.4 (S107) myeloma cell line produces IgA encoded by the same germ-line T15 gene segments used for anti-*P*-Cho antibodies *in vivo*: V_H1, diversity (D) DFL 16.1, and joining (J) J_H1 for the H chain (8) and S107A $V_\kappa22$ and $J_\kappa5$ for the L chain (9). Most of our *in vitro* mutants from S107 have a single amino acid substitution and thus can be used to study the structural and molecular basis of somatic diversification. In addition, S107 has high sequence homology to the *P*-Cho-binding antibody McPC603, whose three-dimensional structure has been determined (10). Here we report a mutant antibody, U10, with a single point mutation[§] in a residue outside and distant from the antigen-binding site that results in complete loss of *P*-Cho binding. The similarity of the McPC603 and S107 antibodies has allowed us to use computational techniques to show how this mutation could cause major steric and electrostatic changes in the antigen-binding site.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Isolation and Purification of U10, S107, and U2U4 Proteins. The U10 and U2U4 mutants were isolated from the S107 cell line, and the proteins were purified by ion-exchange chromatography on a DEAE-Sepharose column (7). S107 protein was purified from ascites fluid by affinity chromatography on *P*-Cho-Sepharose and eluted with 10 mM *P*-Cho (7).

Fluorescence Titrations. Fluorescence titrations were performed as described (11) in the laboratories of Fred Karush (University of Pennsylvania) and Ken Wu (Albert Einstein College of Medicine) with Perkin–Elmer fluorescent spectrophotometers. Proteins were irradiated at 295 nm, and emissions were read at 333 nm.

Direct Binding and Competition ELISAs. Direct binding ELISAs were done as described (12). The polymeric IgA, anti-dextran antibody, W3129, served as a negative control (13). The amount of bound S107, U10, and W3129 proteins was determined by using alkaline phosphatase-linked goat anti-mouse IgA antisera (Zymed Laboratories). Competition ELISAs were performed as in ref. 14. The amount of bound rabbit anti-binding site antibody was detected with enzyme-linked goat anti-rabbit IgG antisera (Zymed Laboratories).

mRNA Sequencing. Preparation and sequencing of U10 and S107 mRNA was done as described (15). H-chain constant (C) region 1 (C_H1)- and V_H-specific primers were used.

Computer Simulation of the ARG-IN and ARG-OUT Models. The program DISCOVER (a product of Biosym Technologies), with force-field parameters as in ref. 16, was used to calculate minimum energy structures and molecular dynamics trajectories. A model of S107 was built from the McPC603 crystal structure (10) by substitution of 32 amino acids in the V_L chain that generally were far from the *P*-Cho-binding site, by substitution of 4 amino acids in the V_H chain, and by insertion of a Tyr residue in H-chain CDR3 between Arg-94 and Asp-101. To substitute amino acids, original atomic side chain positions were matched, and additional atoms were built in an extended conformation. Hydrogen atoms were added geometrically to all crystallographic or built atomic coordinates.

Two U10 models, ARG-OUT and ARG-IN, were built starting from the S107 model with the Asp-101 side chain shortened to an Ala. Models consisted of V_L Asp-1 to Leu-104 and V_H Glu-1 to Thr-110 with N-methyl groups at the carboxyl termini. In the ARG-OUT model, the Arg-94 side chain extended out to the surface as in the McPC603 structure

Abbreviations: *P*-Cho, phosphocholine; V, variable; H, heavy; L, light; J, joining; C, constant; D, diversity.

†Present address: Department of Pediatrics, National Jewish Hospital, Denver, CO 80206.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M24452).

and the S107 model. To build the ARG-IN model, the 13-residue fragment from Tyr-90 to Trp-100I (numbering as in ref. 26, pp. 78 and 204) was extracted, and one terminal nitrogen atom, NH2, of the side chain of Arg-94 was template forced (17) to the position of the *P*-Cho nitrogen atom in the McPC603 crystal structure while the ends of the fragment were fixed in position. The fragment was relaxed by 1.3 psec of dynamics with NH2 still targeted and then incorporated back into the U10 structure to give the starting point for the ARG-IN model.

Water molecules and the sulfate ion were added to complete the U10 models. Crystallographic water molecules within 15 Å of the active site (defined as the sulfate ion and all residues within 4 Å of *P*-Cho; i.e., V_L Phe-91, Tyr-92, Tyr-94, and Leu-96 and V_H Tyr-33, Arg-52, Asp-95, and Trp-100I) were added. From a box of water molecules generated around each complex, water molecules were selected to fill the regions within 7 Å of the active site and the initial Arg-94 position, to fill the regions within 5 Å of charged groups, and to make a thin shell around all antibody residues that had any atom within 15 Å of the active site. Hydrogen atoms were either built to satisfy a clear hydrogen-bonding pattern (crystal waters only) or built randomly to give a H—O—H bond angle of 104.5°.

To gradually relax the U10 models, the nonhydrogen atoms of the antibody were strongly template forced (force constant of 1000.0 kcal·Å^{-1}) to their initial positions during energy minimization (steepest descent algorithm) until the maximum energy gradient was less than 10 kcal·mol^{-1}·Å^{-1}. To complete the minimizations, all waters, the sulfate ion, and antibody residues within 15 Å of the active site were unconstrained while residues outside this area were strongly template forced to their initial positions. Minimization to convergence, necessary for analysis of energetics (18), used the conjugate gradient algorithm with the full force field and no nonbond cutoff. The final root-mean-square energy gradients were less than 0.05 kcal·mol^{-1}·Å^{-1}, and the root-mean-square deviations in position for all nonhydrogen atoms during the last 1000 iterations were about 0.1 Å.

For treatment of the electrostatic potential in the calculation of the energetics of Arg-94, Asp-95, and Glu-35, a macroscopic dielectric constant (ϵ) was used. If all of the electrostatic properties (predominantly electronic polarization and dipoles) of the protein were described microscopically, ϵ would be 1. A dielectric constant of 4 has been derived from studies on crystalline polyamides (19), where both electronic polarization and dipoles are included in the macroscopic constant. Because we explicitly include dipoles as partial point charges on each atom, the major contribution to ϵ should be electronic polarizability, which leads to a value between 2 and 3 (20). Therefore, the energetics in Table 1 were calculated for both $\epsilon = 1$ and $\epsilon = 4$ to bracket the electrostatic potential contribution.

Structural Analysis. To determine the range of conformational variation reasonable for the U10 mutant, the 10 V_L domains and 7 V_H domains superimposed in ref. 21 were compared with the ARG-IN and ARG-OUT models. Computer graphics analysis of all crystallographic and model structures used the programs GRAMPS (22) and GRANNY (23). Electrostatic potential surfaces were calculated to identify critical electrostatic interactions in the *P*-Cho-binding site by using the programs ESPOT and ESSURF (24, 25).

RESULTS AND DISCUSSION

Isolation and Characterization of the U10 Mutant. The U10 mutant was isolated from S107 based on its lack of binding to *P*-Cho (7). Like S107, U10 is a polymeric IgA with two to three H_2L_2 units per molecule and standard IgA interchain disulfide bonding. The H and L chains of U10 were indistinguishable in size from those of S107 by SDS/PAGE, either with or without tunicamycin treatment. However, the U10 antibody did not bind *P*-Cho when assayed by equilibrium dialysis or by enhancement of tryptophan fluorescence (Fig. 1). An ELISA, which is very sensitive for polymeric antibodies like S107, showed no binding of U10 to *P*-Cho, even at high antibody concentrations (Fig. 2).

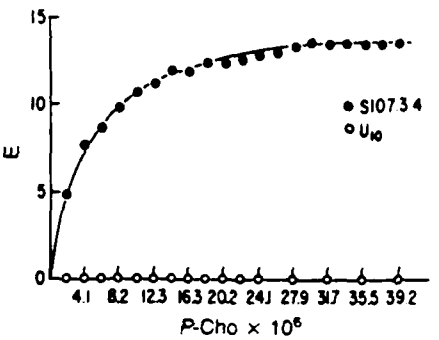

FIG. 1. Comparison of tryptophan fluorescence enhancement of U10 and S107 protein with *P*-Cho. Values for fluorescence enhancement (E) of S107 (●) compared with U10 (○) are calculated and corrected for dilution for 20 points on the experimental curve derived in ref. 11.

A Single Amino Acid Change in the U10 V_H Region. To determine the structural basis for the loss of antigen binding by the U10 antibody, we sequenced and compared the S107 and U10 immunoglobulin V_L and V_H regions and the contiguous C_H1 and C_κ regions directly from mRNA. The S107 and U10 V_H and D region sequences were identical, but the J_H1 region of U10 had a single adenosine to cytidine transversion (Fig. 3), which resulted in the substitution of an Ala residue for Asp-101 (numbering as in ref. 26, pp. 78 and 204). This mutation was confirmed by cDNA sequencing of the U10 H-chain V, D, and J regions. Since the three-dimensional structure of the related McPC603 antibody suggested that residue 101 was not located in the antigen-binding site, we looked for other changes in the nucleotide sequence of the U10 antibody gene. The C_H1 region of the U10 H chain (Fig. 3) was found to be identical with the germ-line S107 sequence (27). The S107 cell line produces two L chains (S107A and S107B), but only the S107A L chain is assembled and secreted with the H chain in the S107 parent and the U10 mutant (9). The V_L, $J_\kappa5$, and C_κ regions of the U10 and S107A L chains were identical, consistent with chain-recombination

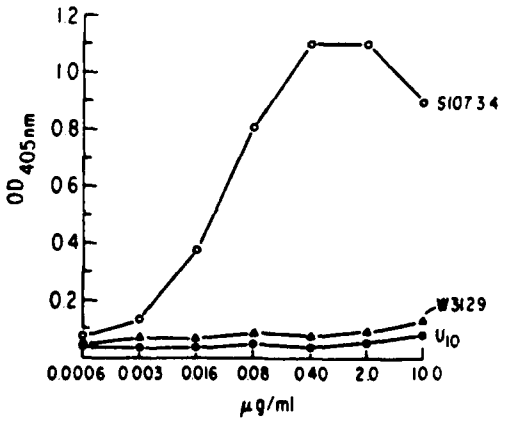

FIG. 2. Direct-binding ELISA of U10 and S107 protein to *P*-Cho. Increasing concentrations of S107 (○), U10 (●), and the non-*P*-Cho-binding control W3129 (△) antibodies were incubated on *P*-Cho-bovine serum albumin-coated ELISA plates. Binding was measured by using subclass-specific goat anti-mouse immunoglobulin linked to alkaline phosphatase and is expressed as OD at 405 nm.

FIG. 3. Sequence comparison of the rearranged H-chain genes of U10 and S107. The nucleic acid sequence includes a portion of the 5' untranslated region and the leader. The V_H, D, J_H, and C_H1 domains of the U10 and S107 genes are shown labeled at the first base of each region. The solid line representing identity of the two sequences is interrupted by the altered codon (in parentheses) and the resulting amino acid substitution in U10. The first (HV1) and second (HV2) hypervariable regions and the *Taq* I site that is lost due to the base change are also shown.

experiments, suggesting that the loss of antigen binding is associated with the H chain (28). Thus, the only sequence difference between the S107 and U10 V and adjoining C region domains is the substitution of an uncharged Ala for negatively charged Asp-101 in the fifth residue of J_H1.

Significant Conformational Change in the U10 Binding Site. To examine the impact of the U10 mutation on V region structure, we examined the reactivity of a number of anti-idiotypic antibodies with the S107 and U10 antibodies. Although eight different anti-T15 monoclonal antibodies, specific for nonbinding site surfaces of the V domain (29), did not distinguish between the U10 and S107 proteins, polyclonal antibodies to the antigen-binding site of S107 did (Fig. 4). Binding of a polyclonal antiserum to S107 could be inhibited by the addition of either S107 or U2U4, a single-site mutant of S107 exhibiting a significant decrease but not a complete loss of *P*-Cho binding (30). In contrast, much higher concentrations of the U10 protein only partially inhibited S107 binding (Fig. 4). The same results were obtained with three different monoclonal anti-binding site antibodies, suggesting a conformational change in the U10 binding site.

Computer Modeling of U10. The three-dimensional structure of the V region of the U10 antibody was analyzed starting from the two crystallographic structures of the S107-like antibody McPC603 (10, 31). One had *P*-Cho bound and the other a sulfate ion. A model of the S107 V domain was built from the sulfate-bound McPC603 structure (Fig. 5*A*) by incorporating appropriate amino acid substitutions. Calculations of the electrostatic potential on the molecular surface of the hapten and the V regions of the McPC603 structure and the S107 model reveal strong electrostatic complementarity between the bound *P*-Cho hapten and the antibodies. These results quantitate local electrostatic stabilization of the hapten suggested in analyses of the crystal structure (10). Residues contributing most to the local electrostatic complementarity with *P*-Cho were all in V_H: Arg-52 with the phosphate moiety and Glu-35 and Asp-95 with the positively charged trimethylammonium group (see Fig. 5*A*).

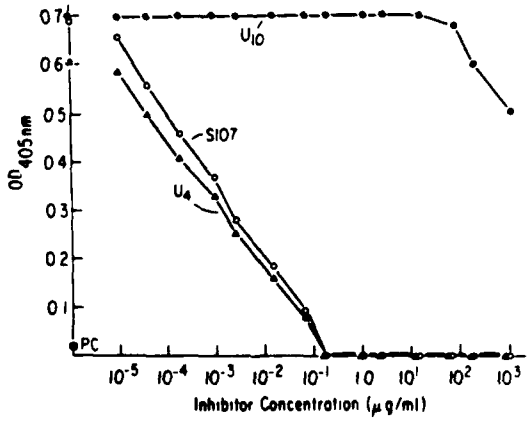

FIG. 4. Determination of a conformational change in the binding site of U10 by competition ELISA. The ability of U10 (●), S107 (○), and U2U4 (▲) proteins to inhibit the binding of a binding-site-specific polyclonal antiserum to S107 protein was tested. Binding is expressed in OD units. At the concentrations used, the polyclonal antiserum is inhibited 100% by *P*-Cho (■) as indicated.

In the McPC603 structures, a hydrogen-bonded salt bridge in V_H between the side chains of Asp-101 and Arg-94 forms at the surface of the antibody, with both side chains extending away from the active site, as seen in the model of S107 in Fig. 5*A*. Asp-101 is 11.8 Å distant from *P*-Cho, and Arg-94 is 9.3 Å distant. Disruption of this salt bridge due to the Ala-101 mutation of U10 could affect the conformation of Arg-94 such that the Arg-94 side chain could extend into the negatively charged pocket that binds the positively charged trimethylammonium moiety of *P*-Cho in the nonmutated antibody. Rotation of the whole length of the Arg side chain by 180° would allow the terminal nitrogen atoms to extend to within 2 Å of the *P*-Cho nitrogen in the bound complex.

To explore this hypothesis, two models of the V region of the U10 protein, termed ARG-IN and ARG-OUT, were built from the S107 model by substitution of Asp-101 with Ala by side-chain truncation. Computer graphics followed by a combination of template forcing, molecular dynamics, and full energy minimization, including explicit solvent molecules, yielded stereochemically reasonable structures. In the ARG-OUT model, the side-chain positions were maintained as in the McPC603 structure, while in the ARG-IN model, the Arg-94 side chain was moved into the *P*-Cho-binding site (Fig. 5*B*). Since Arg-94 and Asp-95 are adjacent on a β-strand, bringing Arg-94 into the binding pocket caused the Asp-95 side chain to extend out to the surface away from the binding pocket.

Root-mean-square deviations of both ARG-OUT and ARG-IN from McPC603 framework main-chain positions were well within the range observed in immunoglobulin crystal structures (21), suggesting that main-chain conformational changes made in the ARG-IN model to accommodate the Arg-94 side-chain repositioning are plausible. Superposition of individual immunoglobulin V_L and V_H domains onto

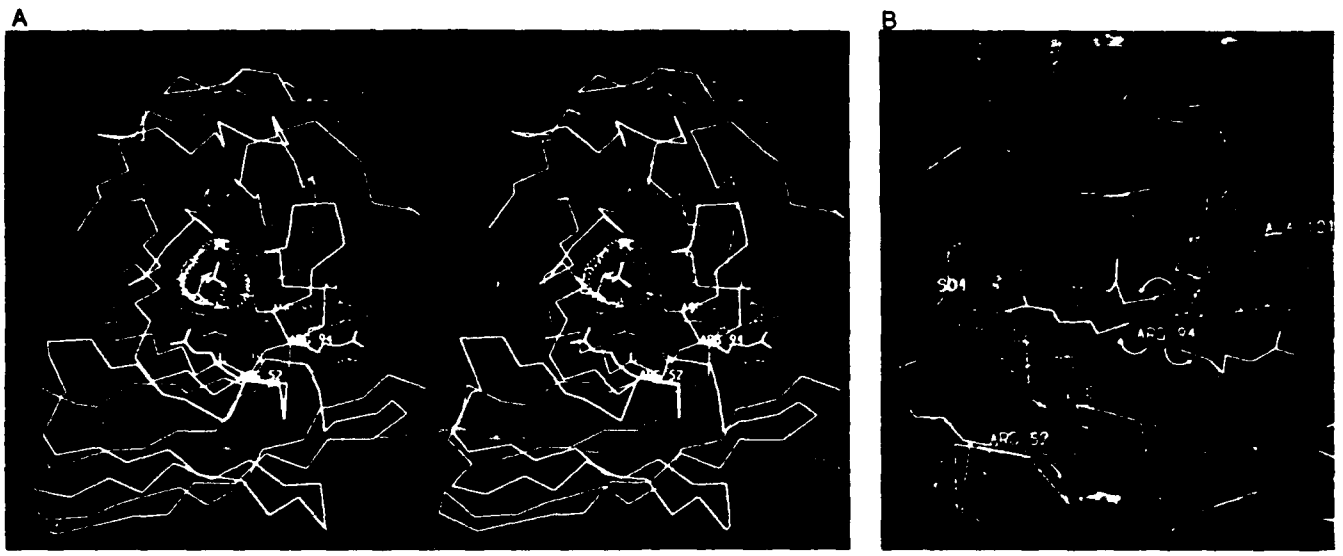

FIG. 5. Computer graphics of the derived S107 and U10 models. (*A*) A stereo pair showing the location and stereochemistry of critical charged-residue side chains in the computationally derived model of the *P*-Cho-binding antibody S107. The α-carbon backbones for the H (blue, bottom) and L (light purple, top) chains are shown together with residue labels and bonds for positive (light blue) and negative (red) side chains and *P*-Cho (center) moieties. The trimethylammonium portion of *P*-Cho is shown in light blue (rear center), and the phosphate portion is shown in red (front center). The salt bridge between Asp-101 (red label, bonds, and surface) and Arg-94 (light-blue label, bonds, and surface) occurs on the outside surface of the antibody (right) over 9 Å from the active *P*-Cho-binding site formed at the interface of the H and L chains. (*B*) The minimized computationally derived models for the ARG-IN and ARG-OUT mutant structures. The α-carbon backbones for the ARG-IN (H chain, light blue, bottom; L chain, pink, top) and the ARG-OUT models (H chain, light green, bottom; L chain, purple, top) superimpose closely except for V_H CDR3. The sulfate (yellow label, bonds, and surface), binding-pocket side chains of Arg-52 (blue label and bonds; lower left) and Glu-35 (red label, bonds, and surface; lower center), and the single site mutation at Ala-101 (light-blue label and bond; right) are shown in the energetically favored ARG-IN model only. The side chains of Arg-94 and Asp-95 are shown in both the ARG-IN and ARG-OUT models. In the ARG-OUT model (yellow bonds and arrows from label), Arg-94 extends out from the *P*-Cho-binding site but has lost its charged interaction with residue 101 (Ala in U10), while Asp-95 points into the *P*-Cho-binding site containing negatively charged Glu-35. In the ARG-IN model, Arg-94 (light-blue bonds, label, and arrow) extends into the binding site, and Asp-95 (red bonds, label, and arrow) extends to the outside surface and points away from the *P*-Cho-binding site.

the corresponding domains of McPC603 gave root-mean-square deviations varying from 0.48 Å to 1.69 Å for V_L and from 0.80 Å to 1.29 Å for V_H, with values for ARG-IN of 0.48 Å for V_L and 0.72 Å for V_H and for ARG-OUT of 0.48 Å for V_L and 0.56 Å for V_H. Moreover, computer graphic examination of 10 superimposed V_L domains and 7 V_H domains indicated that H-chain sites in HyHEL5 (residue 8), HEd10 (residue 44), and Newm (residue 73) as well as the L-chain site in Kol (residue 12) showed greater framework variability than in our ARG-IN model of U10 at H-chain residues 94–95.

Role of Arg-94 in Binding Specificity. To understand the differences between the ARG-IN and ARG-OUT models in more detail, energetics for three critical combining site residues (Arg-94, Asp-95, and Glu-35 of V_H) were calculated (Table 1). To bracket the electrostatic contribution, the energetics were calculated with a dielectric constant (ε) of both 1 (probably overestimates the electrostatic contribution) and 4 (underestimates). The energetics of Arg-94 (Table 1), more favorable by 1.4 kcal·mol^{-1} (ε = 4) to 4.0 kcal·mol^{-1} (ε = 1) for the ARG-IN model compared with ARG-OUT, reflect that the water-solvated Arg-94 side chain of the McPC603-like model ARG-OUT extends away from the *P*-Cho-binding site with no formation of new intramolecular hydrogen-bonding interactions, whereas the Arg-94 side chain of ARG-IN extends into the active site and forms two hydrogen-bonded salt bridges, one with the sulfate ion and one with V_H Glu-35 (Fig. 5*B*). The energetics of Asp-95 are much more favorable in the ARG-IN model by 32.6 kcal·mol^{-1} for ε = 1, while the energetics are slightly less favorable by 0.8 kcal·mol^{-1} for ε = 4. Thus, the ARG-IN model indicates that there is increased stability of the U10 mutant when the Arg-94 side chain is positioned in the negative potential of the binding-site pocket while the Asp-95 side chain extends out and away from the pocket into solvent.

Table 1. Energetics of H-chain Arg-94, Asp-95, and Glu-35 for the two U10 mutant models ARG-IN and ARG-OUT

	Potential energy, kcal·mol^{-1}					
	ε = 1			ε = 4		
	Arg-94	Asp-95	Glu-35	Arg-94	Asp-95	Glu-35
			ARG-IN			
V_L	48.3	−66.5	−67.9	7.7	−17.6	−18.1
V_H	−109.7	−78.2	−142.4	−23.7	−18.6	−34.9
Sulfate	−115.4	43.8	71.5	−29.3	11.0	17.9
Waters	46.4	−93.2	−38.1	11.0	−19.4	−9.8
Total	−130.4	−194.1	−176.9	−34.3	−44.6	−44.9
			ARG-OUT			
V_L	57.9	−61.5	−74.4	13.9	−17.9	−20.6
V_H	−119.8	−96.0	−93.7	−31.5	−27.6	−27.6
Sulfate	−38.1	66.0	64.0	−9.5	16.5	16.0
Waters	−26.4	−70.0	−52.5	−5.8	−16.4	−13.5
Total	−126.4	−161.5	−156.6	−32.9	−45.4	−45.7

To bracket the electrostatic contribution, potential energies were calculated for both a dielectric of 1 (ε = 1) and a dielectric of 4 (ε = 4). The potential energy of each of these residues with the V_L chain, the sulfate ion, and the waters is due to nonbond interactions, the sum of electrostatic and van der Waals energies. For example, the interaction between Arg-94 and the V_L chain is the sum of the interactions between each atom of Arg-94 and each atom of the V_L chain. The potential energy of each residue with the V_H chain is the sum of the nonbond energies with all residues in the V_H chain plus the internal energy (strain in the bond, angle, and torsion angle energies) calculated for that specific residue.

In the ARG-IN model, the V_H Glu-35 side chain changes conformation to allow salt-bridge formation with the Arg-94 side chain (shown as intersecting surfaces in Fig. 5*B*). Glu-35 forms this salt bridge while maint ng its hydrogen bond to the side chain of L-chain Tyr-9 u-35 does not form any salt bridges in the crystal stru e of McPC603 or in the minimized ARG-OUT structure. For Glu-35 (Table 1), the salt-bridged ARG-IN model is energetically more favorable (by 20.3 kcal·mol^{-1}) for $\epsilon = 1$, whereas for $\epsilon = 4$, it is slightly less favorable (by 0.8 kcal·mol^{-1}). Thus, having the Arg-94 side chain in the active site puts Glu-35, a residue known to be important in *P*-Cho binding (30, 32), in a more energetically favorable position.

Among related antibodies (ref. 26, pp. 204–208) with sequence changes for Asp-101, several in subgroup IIIA have replacements of Asp-101 (to Ala or Gly), while retaining Arg-94. Only one of these (CBA/J 6G6) is known to bind *P*-Cho. If, as we have explored above, the Asp-101 residue is necessary to form a salt bridge with Arg-94 so that *P*-Cho binding is retained, how can we explain this exception? CBA/J 6G6 has a much shortened CDR3 between Arg-94 and Ala-101 and has Asp-98 two residues before Ala-101, suggesting that Asp-98 can replace Asp-101 in the salt bridge to Arg-94.

Role for Side-Chain Rearrangements in Antibody Diversity. We have found that a single base pair mutation in the J_H1 gene of the S107 V_H region, corresponding to an amino acid outside and distant from the active site, completely eliminates antigen binding in the somatic mutant U10. Similar observations have been reported in anti-lysozyme antibodies (33). Our results and observations by others of substitutions in J_H1 (1, 34) and in J_H3 and J_H4 (35) suggest that residues distant from the binding site may play an important role in the specificity and affinity of the antigen-binding site. The loss of *P*-Cho binding and serological studies with binding-site-specific antibodies indicate a significant conformational change local to the antigen-binding site in U10 greater than that created by the Glu-35 to Ala substitution in the binding site of the single-site S107 mutant U2U4.

Our two models of the U10 antibody, ARG-IN and ARG-OUT, differ in placement of the Arg-94 side chain. In the ARG-IN model, the Arg guanidinium group lies in the position that would be occupied by the trimethylammonium group of *P*-Cho and results in more favorable energies for Arg-94, Asp-95, and Glu-35 compared with the ARG-OUT model (Table 1). Thus, a role of the Asp-101 side chain in McPC603 and S107 may be to keep the Arg-94 side chain out of the negatively charged binding site by forming a salt bridge to it on the exterior of the antibody. Lack of the Arg-94 to Asp-101 salt bridge in the U10 mutant may allow the Arg-94 side chain to be folded into the active site, as modeled by ARG-IN, explaining the complete loss of *P*-Cho binding with little overall conformational change.

Energy minimizations were done at the San Diego Supercomputer Center. We thank Dr. Wendy Cook for her tryptophan fluorescence data depicted in Fig. 1 and John Tainer, Michael Pique, Terri Kelly, Terry Potter, and Rosalie Spata for critical advice and help in preparing the manuscript. This work was supported by National Institutes of Health Grants CA-39838, AI-05231, and AI-10702 to M.D.S. and Office of Naval Research Grant N00014-89-J-1174 to E.D.G. N.C.C. and A.M.G. were supported by a training grant from the National Cancer Institute (CA-09173). M.D.S. is supported in part by the Harry Eagle chair in Cancer Research from the National Women's Division, Albert Einstein College of Medicine.

1. Clarke, S. H., Huppi, K., Ruezinsky, D., Staudt, L., Gerhard, W. & Weigert, M. (1985) *J. Exp. Med.* **161**, 687–704.
2. Allen, D., Cumano, A., Dildrop, R., Kocks, C., Rajewsky, K., Rajewsky, N., Roes, J., Sablitzky, F. & Siekevitz, M. (1987) *Immunol. Rev.* **96**, 5–22.
3. Berek, C. & Milstein, C. (1987) *Immunol. Rev.* **96**, 23–41.
4. Malipiero, U. V., Levy, N. S. & Gearhart, P. J. (1987) *Immunol. Rev.* **96**, 59–74.
5. Manser, T., Wysocki, L. J., Margolies, M. N. & Gefter, M. L. (1987) *Immunol. Rev.* **96**, 141–162.
6. Claflin, J. L., Berry, J., Flaherty, D. & Dunnick, W. (1987) *J. Immunol.* **138**, 3060–3068.
7. Cook, W. D. & Scharff, M. D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5687–5691.
8. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25**, 59–66.
9. Kwan, S. P., Max, E. E., Seidman, J. G., Leder, P. & Scharff, M. D. (1981) *Cell* **26**, 57–66.
10. Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986) *J. Mol. Biol.* **190**, 593–604.
11. Jolley, M. E. & Glaudemans, C. P. J. (1974) *Carbohydrate Res.* **33**, 377–382.
12. Spira, G., Bargellesi, A., Pollock, R. R., Aguila, H. L. & Scharff, M. D. (1985) *Hybridoma Technology in the Biosciences and Medicine* (Plenum, New York), pp. 77–78.
13. Barstad, P., Farnsworth, V., Weigert, M., Cohn, M. & Hood, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4096–4100.
14. Chien, N. C., Pollock, R. R., Desaymard, C. & Scharff, M. D. (1986) *J. Exp. Med.* **167**, 954–973.
15. Geliebter, J., Zeff, R. A., Schulze, D. H., Pease, L. R., Weiss, E. H., Mellor, A. L., Flavell, R. A. & Nathenson, S. G. (1986) *Mol. Cell Biol.* **6**, 645–652.
16. Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M. & Hagler, A. T. (1988) *Proteins* **4**, 31–47.
17. Struthers, R. S., Rivier, J. & Hagler, A. T. (1985) *Ann. N.Y. Acad. Sci.* **439**, 81–96.
18. Roberts, V. A., Dauber-Osguthorpe, P., Osguthorpe, D. J., Levin, E. & Hagler, A. T. (1986) *Isr. J. Chem.* **27**, 198–210.
19. Baker, W. O. & Yager, W. A. (1942) *J. Am. Chem. Soc.* **64**, 2171–2177.
20. Gilson, M. K., Rashin, A., Fine, R. & Honig, B. (1985) *J. Mol Biol.* **184**, 503–516.
21. Getzoff, E. D., Tainer, J. A., Lerner, R. A. & Geysen, H. M. (1988) *Adv. Immunol.* **43**, 1–98.
22. O'Donnell, T. J. & Olson, A. J. (1981) *Comput. Graph.* **15**, 133–142.
23. Connolly, M. L. & Olson, A. J. (1985) *Comput. Chem.* **9**, 1–6.
24. Getzoff, E. D., Tainer, J. A., Weiner, P. K., Kollman, P. A., Richardson, J. S. & Richardson, D. C. (1983) *Nature (London)* **306**, 287–290.
25. Getzoff, E. D., Tainer, J. A. & Olson, A. J. (1986) *Biophys. J.* **49**, 191–206.
26. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (Dept. Health Human Serv., Washington, DC).
27. Tucker, P. W., Slightom, J. L. & Blattner, F. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7684–7688.
28. Giusti, A. M. (1984) Dissertation (Albert Einstein Coll. Med., Yeshiva Univ., Bronx, NY).
29. Desaymard, C., Giusti, A. M. & Scharff, M. D. (1984) *Mol. Immunol.* **21**, 961–967.
30. Giusti, A. M., Chien, N. C., Zack, D. J., Shin, S. & Scharff, M. D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2926–2930.
31. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
32. Padlan, E. A., Cohen, G. H. & Davies, D. R. (1985) *Ann. Inst. Pasteur/Immunol.* **136C**, 259–263.
33. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljack, R. J. (1986) *Science* **233**, 747–753.
34. Gough, N. M. & Bernard, O. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 509–513.
35. Smith-Gill, S. J. (1985) in *Current Communications in Molecular Biology: Immune Recognition of Protein Antigens*, eds. Laver, W. G. & Air, G. M. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 61–65.